

A Convex Functional for Image Denoising based on Patches with Constrained Overlaps and its vectorial application to Low Dose Differential Phase Tomography

Alessandro Mirone

*European Synchrotron Radiation Facility,
BP 220, F-38043 Grenoble Cedex, France*

Emmanuel Brun

*European Synchrotron Radiation Facility and the Ludwig Maximilians University,
Physics Departement, 85748 Garching, Germany*

Paola Coan

*Ludwig Maximilians University, Departement of Clinical Radiology,
80336 Munich, Germany and LMU,
Physics Departement, 85748 Garching, Germany*

Abstract

We solve the image denoising problem with a dictionary learning technique by writing a convex functional of a new form. This functional contains beside the usual sparsity inducing term and fidelity term, a new term which induces similarity between overlapping patches in the overlap regions. The functional depends on two free regularization parameters: a coefficient multiplying the sparsity-inducing L_1 norm of the patch basis functions coefficients, and a coefficient multiplying the L_2 norm of the differences between patches in the overlapping regions. The solution is found by applying the iterative proximal gradient descent method with FISTA acceleration. In the case of tomography reconstruction we calculate the gradient by applying projection of the solution and its error backprojection at each iterative step. We study the quality of the solution, as a function of the regularization parameters and noise, on synthetic datas for which the solution is a-priori known. We apply the method on experimental data in the case of Differential Phase Tomography. For this case we use an original approach which consists in using vectorial patches, each patch having two components: one per each gradient component. The resulting algorithm, implemented in the ESRF tomography reconstruction code PyHST, results to be robust, efficient, and well adapted to strongly reduce the required dose and the number of projections in medical tomography.

I. INTRODUCTION

Users of X-ray tomography aim to push the frontiers of their studies towards new domains which require finer time resolution, better signal to noise ratio, and less radiation damage. All these three requirements bring to a data-starving situation where for a given quality goal, the available data volume is never enough. A solution to this problem consists in filling in the gap left by the missing data with an a-priori knowledge of the solution. The signals occurring in Nature, when cleaned from noise, present most of the time an intrinsic sparsity when expressed in the proper basis. An image is intrinsically sparse when it can be approximated as a linear combination of a small number n of basis functions with $n \ll N$, where N is the image dimensionality. Piece-wise constant images are examples of sparse signal : they have non-zero signal only at the flat regions borders when they are expressed by their gradient. For piece-wise constant images one can apply very efficient methods based on minimization of a convex functional, said also convex objective function, which contains a total variation penalty term. For other classes of images such as medical images one has to choose different solutions which are adapted to the intrinsic sparsity of the case under study. There are mainly two ways : either the sparsity structure is a-priori known and an appropriate basis of functions can be built from the beginning, or it must be automatically learned from a learning set with the dictionary learning technique. The dictionary learning technique has recently been applied to tomography reconstruction using the Orthogonal Matching Pursuit (OMP) denoising procedure[1]. This procedure consists in obtaining first an over-complete basis of functions and then in fitting every patch of the image using at most N_{omp} components selected from this basis. The components are heuristically selected choosing, each time, the one having the maximum overlap with the remaining error. This optimization method cannot be implemented as a convex objective function optimization problem because the linear combination of two candidate solutions can have more than N_{omp} components. In other words the optimization domain is not convex. In this paper we present a more advanced formalism based on a convex functional that we describe in section II. For the solution of our functional minimization problem we have applied the recently developed tools taken from the field of convex optimization[2]. The result is a robust and efficient algorithm that we discuss and illustrate with synthetic and medical data in section III.

II. METHODS

In this section we introduce first the decomposition of an image into non-overlapping patches and the related objective function for denoising. Then we introduce our original formalism which ensures, using overlapping patches, a smooth transition at the patches borders, and finally we apply this formalism to image denoising and tomography reconstruction.

We denote by $\mathbf{1}_p$ the indicator function of patch p , which is equal to 1 over the patch support (typically an $m \times m$ square) and is zero elsewhere. For non-overlapping patches, covering the whole domain, we have :

$$\sum_p \mathbf{1}_p(i) = 1 \quad \forall i. \quad (1)$$

where i denotes the pixel position and can be thought as a two-dimensional vector. We are looking for the ideal solution \mathbf{x} that we express by the vector, \mathbf{w} , of its coefficients in the basis of patch functions:

$$x_i = \sum_p \mathbf{1}_p(i) \sum_k w_{kp} \varphi_k(i - r_p). \quad (2)$$

where the set $\{\varphi_k\}$ is an over-complete basis of functions over the patch support; r_p is the closest to the origin corner of the patch p , and w_{kp} is the component k, p of vector \mathbf{w} which multiplies the basis function φ_k in the patch p . The denoising problem, given an image \mathbf{y} , consists in finding the minimum of a functional $F(\mathbf{w}) = f(\mathbf{w}) + g(\mathbf{w})$ which is sum of two terms. The term $f(\mathbf{w}) = \|\mathbf{y} - \mathbf{x}\|_2^2$ links the solution to the data \mathbf{y} . The other term, $g(\mathbf{w})$, contains the a-priori knowledge about the solution. This way of breaking the functional in two terms has his roots in the Bayesian theorem. From a probabilistic point of view the denoising problem consists in finding, given a noisy image $\underline{\mathbf{y}}$ of an object, the most probable object \mathbf{x} that can generate that image. We represent the object \mathbf{x} through the patches coefficients \mathbf{w} . The Bayes theorem, applied to denoising, states that the conditional probability of \mathbf{w} being the exact object given a measurement \mathbf{y} , is the product of the probability of \mathbf{y} being the measure given the exact solution \mathbf{w} , times the a-priori probability of \mathbf{w} .

Assuming gaussian noise, the conditional probability of \mathbf{y} being the measure given the exact solution \mathbf{w} is $\exp(-\|\mathbf{y} - \mathbf{x}\|_2^2 / (2\sigma^2))$, where \mathbf{x} is expressed through the patches co-

efficient \mathbf{w} by equation 2. The a-priori probability of \mathbf{w} is written as $\exp(-g(\mathbf{w})/(2\sigma^2))$. The a-priori knowledge that the solution is sparse in our patches basis can be expressed by using the sparsity-inducing L_1 penalization [3]: $g(\mathbf{w}) = \beta \|\mathbf{w}\|_1$. The most probable solution \mathbf{w}^* is obtained by finding the $F(\mathbf{w})$ minimum:

$$\begin{aligned}\mathbf{w}^* &= \operatorname{argmin}_{\mathbf{w}}(f(\mathbf{w}) + g(\mathbf{w})); \\ f(\mathbf{w}) &= \|\mathbf{y} - \mathbf{x}\|_2^2; \quad g(\mathbf{w}) = \beta \|\mathbf{w}\|_1.\end{aligned}\tag{3}$$

The solution can be obtained by using the iterative shrinkage thresholding algorithm[2] (ISTA) iterations:

$$\mathbf{w}_{n+1} = T_{\beta\gamma}(\mathbf{w}_n - \gamma \nabla f(\mathbf{w}_n)); \quad \mathbf{w}^* = \mathbf{w}_\infty.\tag{4}$$

where T_α is the shrinkage operator defined as

$$T_\alpha(\mathbf{w}) = \frac{\mathbf{w}}{\|\mathbf{w}\|_2} \max(\|\mathbf{w}\|_2 - \alpha, 0).\tag{5}$$

and γ is a positive number lesser than the inverse of the Lipschitz condition number L :

$$\gamma \in]0, 1/L].\tag{6}$$

The Lipschitz number L is such that :

$$\|\nabla f(\mathbf{w}_2) - \nabla f(\mathbf{w}_1)\|_2 \leq L \|\mathbf{w}_2 - \mathbf{w}_1\|_2; \quad \forall \mathbf{w}_1, \mathbf{w}_2.\tag{7}$$

The ISTA algorithm can be accelerated by the Fast Iterative Shrinkage Thresholding [4](FISTA) method.

In its non-overlapping version, the image denoising with patches is able to detect features that are within the field of the patch: if a line crosses the central region of a patch, it will be detected if the basis of function has been trained to detect such lines. But in the situation where a line intersects only one point in a corner of the patch square, the signal of this point is indistinguishable from that of a noisy point, no matters the dictionary training.

For this reason the patches denoising technique is often used with overlapping patches using post-process averaging[5]. In this case the minimization problem is solved for each patch separately first, and then the averaging is performed in the overlapped regions.

In this study we do not follow this procedure but we add an overlap term into the objective function. We choose a system of patches which covers the whole domain, and we allow for overlapping. In this case the sum of all indicator functions is greater or equal to one :

$$\sum_p \mathbf{1}_p(i) \geq 1; \forall i. \quad (8)$$

We define the core indicator functions $\mathbf{1}_p^c$, which indicate the core of the patches, and make a non-overlapping covering:

$$\mathbf{1}_p^c(i) \leq \mathbf{1}_p(i); \quad \sum_p \mathbf{1}_p^c(i) = 1; \forall i. \quad (9)$$

For a given point i , $\mathbf{1}_p^c(i)$ indicates which patch p has its center C_p closest to point i :

$$\sum_p \mathbf{1}_p^c(i) \|i - C_p\|_1 \leq \|i - C_{p'}\|_1; \quad \forall p', i. \quad (10)$$

The solution \mathbf{x} is composed using the central part of the patches as indicated by the functions $\mathbf{1}_p^c$:

$$x_i = \sum_p \mathbf{1}_p^c(i) \sum_k w_{kp} \varphi_k(i - r_p). \quad (11)$$

Now we introduce the P operator which is the projection operator, for tomography reconstruction, and is the identity for image denoising. The functional $F(\mathbf{w})$ whose minimum gives the optimal solution is written, for both applications, as:

$$\begin{aligned} F(\mathbf{w}) &= f(\mathbf{w}) + g(\mathbf{w}); \quad g(\mathbf{w}) = \beta \|\mathbf{w}\|_1; \\ f(\mathbf{w}) &= \|\mathbf{y} - \mathbf{P}(\mathbf{x})\|_2^2 + \\ &\quad \rho \sum_{pi} \mathbf{1}_p(i) \left(x_i - \sum_k w_{kp} \varphi_k(i - r_p) \right)^2. \end{aligned} \quad (12)$$

where the ρ factor weights a similarity-inducing term which pushes all the overlapping patches, which touch a point i , toward the value x_i of the global solution x in that point. For future reference we call $\|\mathbf{y} - \mathbf{P}(\mathbf{x})\|_2^2$ the fidelity term.

The solution is found with the FISTA method, using the gradient of $f(\mathbf{w})$ which is easily written in compact form:

$$\begin{aligned}
\frac{\partial f(\mathbf{w})}{\partial w_{kp}} = \sum_i 2\varphi_k(i - r_p) \mathbf{1}_p^c(i) & \left\{ (\mathbf{P}^T (\mathbf{P}(\mathbf{x}) - \mathbf{y}))_i + \right. \\
& \left. \rho \sum_{p'} \mathbf{1}_{p'}(i) \left(x_i - \sum_{k'} w_{k'p'} \varphi_{k'}(i - r_{p'}) \right) \right\} + \\
& \sum_i 2\varphi_k(i - r_p) \rho \mathbf{1}_p(i) \left(\sum_{k'} w_{k'p} \varphi_{k'}(i - r_p) - x_i \right) \quad (13)
\end{aligned}$$

where P^T is the adjoint operator of P , and is called back-projection operator in the case of tomography, and is still the identity for image denoising.

III. APPLICATION

In this section we compare the fit with patches of a given image with or without overlapping. We analyze the quality of the recovered image as a function of the two regularization parameters β and ρ . The same study is then performed on the same image with additive noise. Then we apply the method to tomography reconstruction of virtual phantoms using our convex functional with patch overlapping. We analyze the quality of the reconstruction as a function of the two regularization parameters projections. Finally, we apply the method onto experimental cases for the reduction of number of projection in order to reduce the deposited dose during a tomography.

A. Fitting problem

Figure 1 shows a training image (Lena) on the left. From this image we have extracted by dictionary learning the basis of patches that we show on the right. To obtain the basis we have implemented the K-SVD algorithm[6] that we have applied using four atoms in the OMP procedure. Our basis consists in an over-complete set of 100 functions having a 7×7 pixels support.

We apply this set of patches that we have learned from Lena on a completely different image (Boy [7]). In figure 2 we show the effect of fitting with non-overlapping patches a noiseless image. The original image is subfigure 2a). It has been normalized to have its



FIG. 1: A training image and its K-SVD basis

maximum equal to 1 and its minimum equal to 0. From left to right, of the first row, in the three columns on the right, we have the results for beta equal to 0.2, 0.02, 0.002. The sparsity s is the ratio of non-zero components over the total number of components (100 components per every patch here). The value of sparsity for each image is reported on its legend. When we take the sparsity average over all the image, it takes the values 0.029, 0.13, 0.24 when beta goes from 0.2 to 0.002 in the first row of our example.

From figure 2 it is obvious how the fit with non overlapping patches gives discontinuity at the patches borders. The upper-right corners of the figures are zooms in the hairs zone. We can see that, in the case of non-overlapping patches, features in form of lines are more sensitive to the regularization term when they are close to the patch border than when they are close to the center. When the regularization parameter is weakened by lowering beta, then more components are available for the fit and the dimensionality gets high enough to fit all the features.

The second row of figure 2 shows the same fitting problem solved with our overlapping patches method. The set of overlapping patches has been obtained by replicating the original set of non-overlapping patches by translation vectors $(3l, 3n)$; where l and n are integers in the interval $[0, 2]$. This choice of translation vectors results in a nine-fold overlap : each point is covered by nine patches. The β values take, in the last three columns of the second row, the values 0.09, 0.009, 0.0009 from left to right. The overlap constraining factor is $\rho = 1$.



FIG. 2: Fitting with no-overlapping (first row) and overlapping (second row) patches a noiseless image. The result for $\beta = 0.2/9$ (e) reproduces the same fidelity/ L_1 ratio obtained with $\beta = 0.2$ (b) in the non-overlapping case. For the same level of fitted details the overlapping method requires a higher number of components (bigger s) in order to satisfy the overlapping constraints.

With these values these images have nearly the same sparsity than the non-overlapping results in the same columns of the first row.

One can see in figure 2 that using overlapping patches the discontinuities have disappeared and the lines are homogeneously represented along their length even for the strongly regularized case. The overlap constraint reduces the tessellation effects but we can see that for the same sparsity ratio, another kind of distortion, in the form of smoothing, appears. This is due to the fact that the fidelity term, which links the data to the solution, is proportional to the area of the image, while the L_1 term acts on all the w_{pk} coefficients whose number increases not only with the area of the image but also with the number of replica. Taking into account that we have 9 replica per patch, we show in the first image of second column the result obtained with $\beta = 0.2/9$. This rescale back the L_1 term to a situation which is comparable with the $\beta = 0.2$ case of image b) of the first row. The level of preserved details is comparable in these two cases. The sparsity, however is bigger in the overlapping case: $s = 0.068$ which means nearly seven functions per patch. In the non-overlapping case, instead, it was about 3 components per patch. This means that the similarity inducing

term is constraining a part of the available degrees to realize smooth transitions between neighboring patches.

B. Denoising problem

For the denoising problem two things must be kept in mind to select the regularization weights: for an image with a stronger noise a stronger value of β will be necessary, but it is also true that for a stronger value of the regularizing parameter more features of the noiseless image will be filtered out. The best value of the regularization parameters is therefore a compromise between the necessity of filtering out the noise and that of preserving image features.

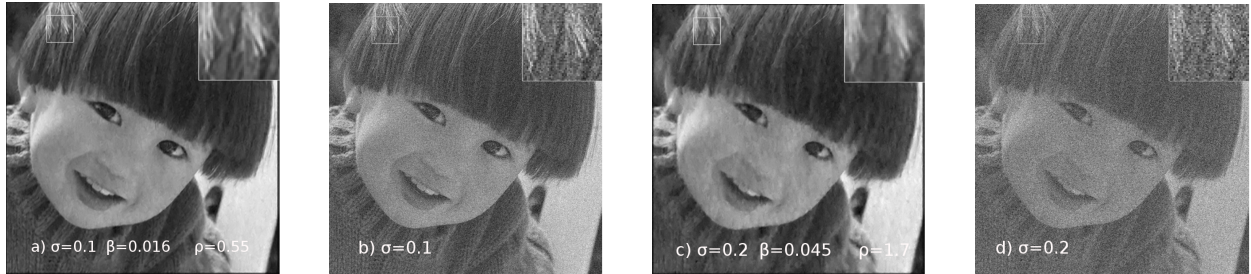


FIG. 3: the effect of overlapping denoising (a,c) on an image with $\sigma = 0.1, 0.2$ (b,d) added white gaussian noise. The denoising has been performed with a replication step of 3 and the optimal β, ρ

We define the quality improvement factor Q based on the the Structural SIMilarity (SSIM [8]) index S which is 1 when the images are identical. Our definition of Q is :

$$Q = (1 - S(\mathbf{y}_{noised}, \mathbf{y})) / (1 - S(\mathbf{y}_{denoised}, \mathbf{y})) \quad (14)$$

We show in figure 4 the quality improvement factor Q for different values of β, ρ for the overlapping case and the non-overlapping one. The tests have been done adding a white gaussian noise ($\sigma_{\{noise\}} = 0.1$ and 0.2) to the original image of figure 2. For the overlapping case a replication step of 3 has been used and we have varied separately β (squares) and ρ (circles) around the optimal values. In the non-overlapping case (stars) only β is varied.

We observe that the quality improvement factor, for the overlapping case, peaks at a β value which is about 1/9 of the β value of the no-overlapping peak. As explained before

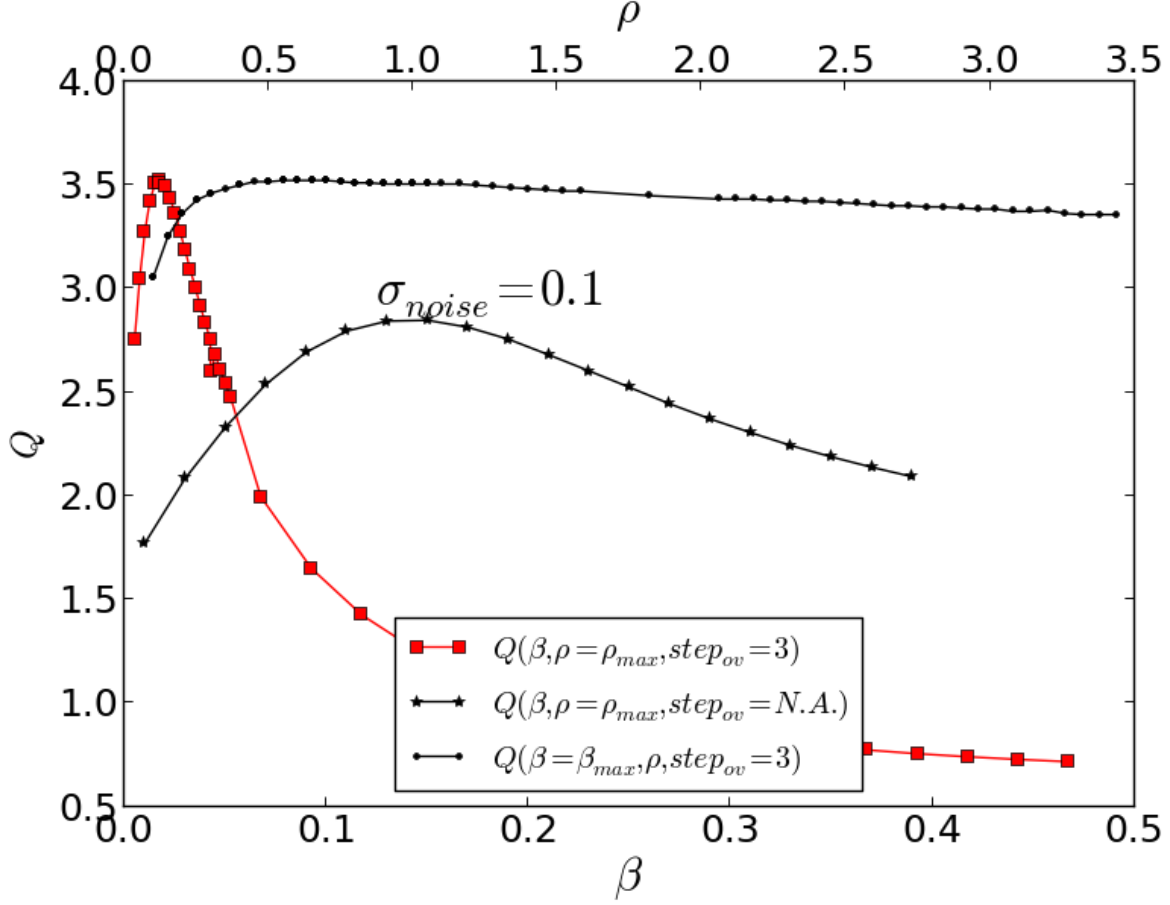


FIG. 4: The quality improvement factor Q for different values of β, ρ around the optimal values for $\sigma_{\{noise\}} = 0.1$. For the overlapping case a replication step of 3 has been used and we have varied separately β (squares) and ρ (circles) around the optimal values. For the non-overlapping case only β is varied.

this lower β is compensated by the nine-fold increase in the number of components which are weighted by L_1 norm. We note that the optimal quality is better with the overlapping method. The trend of β as a function of the noise level is as expected : a stronger noise needs to be regularized with a stronger β .

In figure 3 we show the denoising results (a and c) for two noised images \mathbf{y}_{noised} (b and d) obtained adding a $\sigma_{\{noise\}} = 0.1, 0.2$ strong white gaussian to the original image : $\mathbf{y}_{noised} = \mathbf{y}_{original} + noise(\sigma_{\{noise\}})$. The denoising has been performed with a replication step of 3 pixels and with the β and ρ giving the optimal SSIM.

C. Dose reduction in computed tomography

In this section we apply the method using overlapping patches for tomography reconstruction on synthetic data. The figure 5 presents a 1024×1024 pixels phantom and a bases of patches that has been learned from the phantom .

We show in figure 6 the reconstruction obtained from 150 projections between 0 and 180 degree. Note that this is well below the number of projections, of the order of thousand, that would be required in this case by the Shannon-Nyquist sampling theorem. The lack of information due to the small number of projections results in a noised reconstruction when the filtered-back-projection reconstruction is applied (left) while the image recovered with our method (right), with $(3l, 3n)$ translations for replication, maintains a high quality.

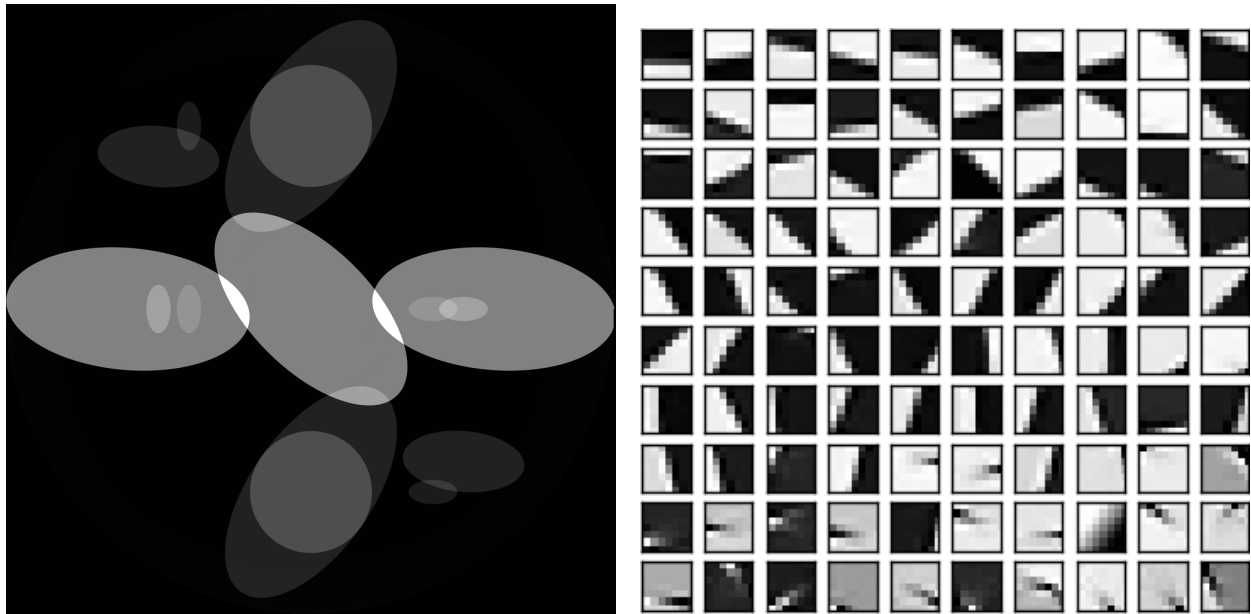


FIG. 5: A phantom and its K-SVD basis

For this noiseless case the choice for β and ρ was easy and figure 6 was produced with our first guess which gave already an excellent reconstruction, without need of further optimization. The first guess was $\beta = 0.001$ and $\rho = 50$. The quality improvement Q between the FBP result and our method is 20. The improvement factor is in this case the FBP error (the SSIM index distance from 1 as in 14) divided by the error of our method.

The noisy case is instead more complex. To study it we add to the phantom sinogram a gaussian white noise with a σ equal to 5% of the maximum sinogram value. In figure 7 we

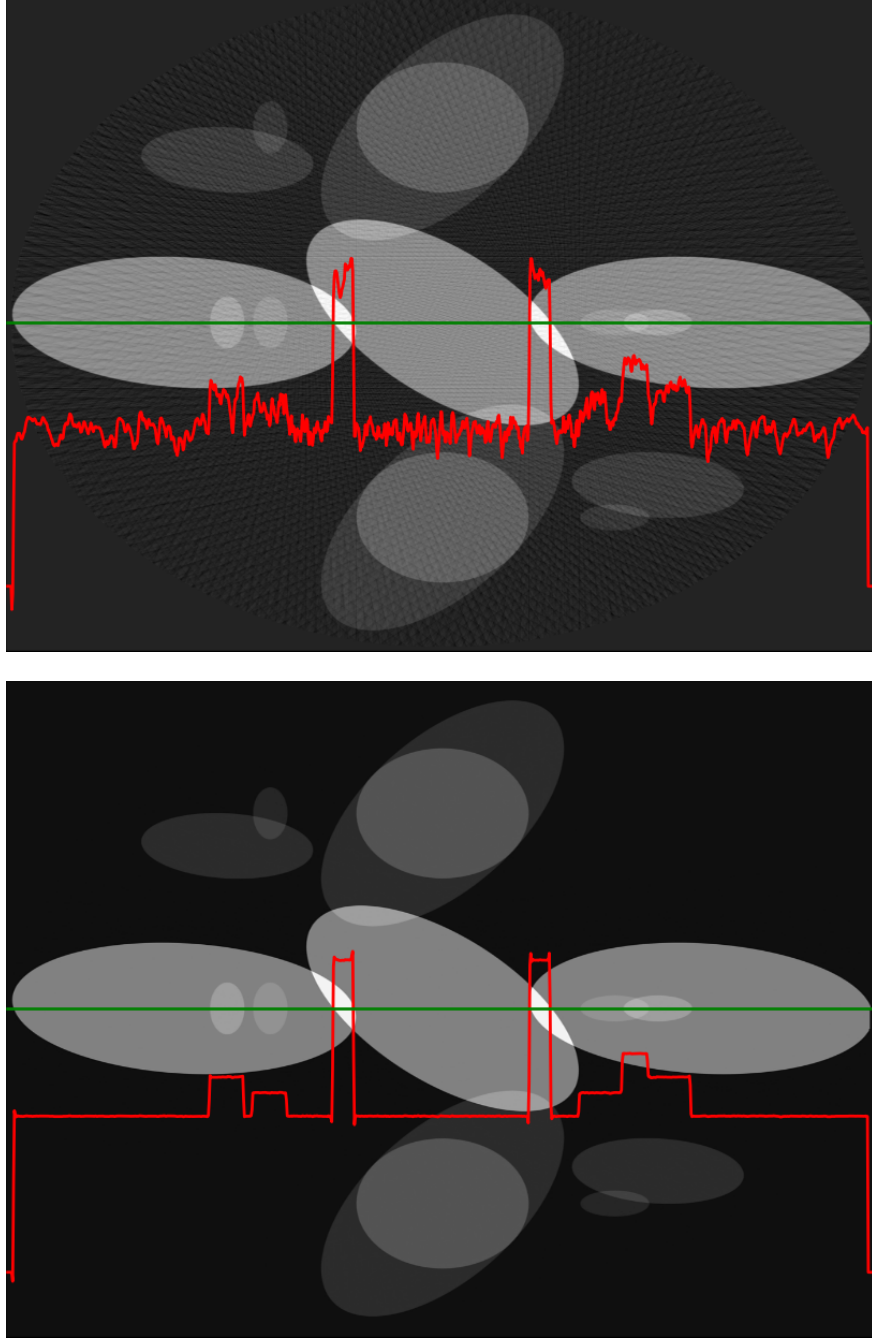


FIG. 6: reconstruction from 150 projections of a 1024×1024 phantom. Top: FBP; Down: our method with $\beta = 0.001$ and $\rho = 50$.

show the Q dependency versus β (squares) and ρ (circles) not far from the optimal choice. We have not fully optimized Q because we have found that in this extreme case where ρ tends to be very big, we get a substantial decrease in convergence speed after the shoulder in

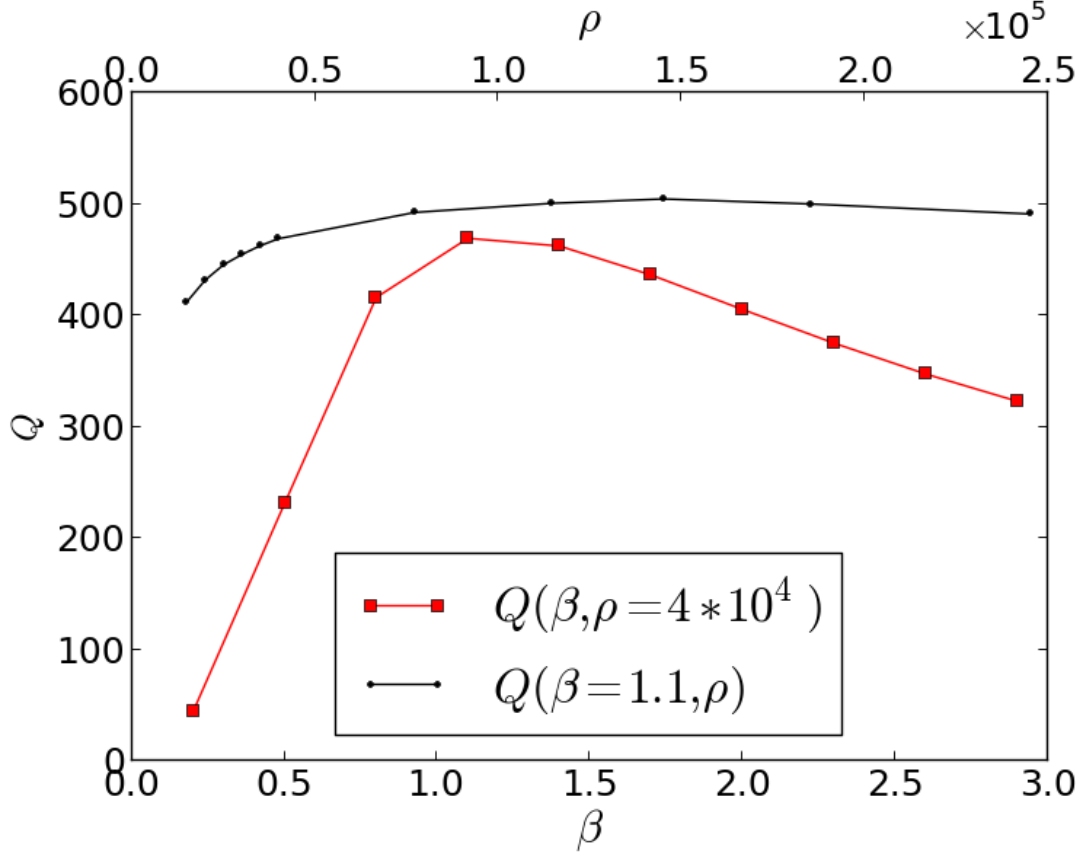


FIG. 7: the Q dependency versus β (squares) and ρ (circles) not far from the optimal choice of β and ρ for the reconstruction of the phantom from 150 noisy projection.

the ρ (circles) curve. Before the shoulder we have convergence in some hundreds of iterations while after the shoulder it takes some thousands to converge.

We observe that at these high values of ρ the effective behavior of our method is similar to those of a total variation penalization : the image is flattened over large regions. We think that the increase in convergence time is due to the fact that the algorithm takes a longer time to propagate when the flattened regions are larger.

Figure 8, bottom, shows the result obtained with our method using the $\beta = 1.1$ and $\rho = 2.5 \times 10^5$ values obtained from the Q optimization illustrated figure 7. The FBP reconstruction is reported at the top of the figure. On these images it is clear that that our method outperforms standard FBP. The ellipsoids become visible and the plotted profile follows the expected behavior, whilst the FBP reconstruction shows only noise.

D. Differential Phase tomography

Finally we show a promising application of our method to medical tomography for a sample imaged using X-ray phase contrast imaging(PCI). This technique has shown an enhancement of soft tissue visualization in comparison to conventional techniques [9]. PCI employs the dual property of X-rays of being simultaneously absorbed and refracted while passing through a tissue. Among all the phase contrast techniques we chose to test our method on the analyzer based imaging because of the high sensitivity and unique results provided by this modality for investigating large and highly absorbing biological tissues (i.e. full human breasts) [10]. Because breast are highly radio sensitive organs, X-ray CT of these organs are not clinically applied, even if a 3D would be benefit for radiologists. A reduction of the deposited radiation dose in CT combined with the unprecedented contrast improvement offered by PCI is thus of high interest for breast cancer detection.

In the analyzer based PCI technique, the projection data contain a signal which is proportional to the gradient of the X-ray phase in one direction (i.e. the direction perpendicular to the plane formed by the incoming and diffracted X-rays on a perfect Bragg crystal which is used for analyzing the radiation passing through the sample). When the object is rotated around an axis (Z-axis, for instance), this signal contains contributions from the X and Y gradient components, where the X and Y axis corotate with the sample. The two components are de-phased by a rotation angle of 90 degrees and can be reconstructed separately by multiplying before-hand the sinogram with the cosine and sine of the rotation angle. We apply our formalism considering that the reconstructed and learning images are vectorial objects : the value associated with a pixel is not a scalar but a two-component vector. More details on the principles and technical aspects of PCI are available in [9].

The sample studied is a 7cm human breast imaged with a pixel size of 100 μm . The experiment was conducted at the biomedical beam line at the European Synchrotron Radiation Facility (ESRF). The sample was a human breast mastectomy specimen. The study was performed in accordance with the Declaration of Helsinki and was approved by the local ethics committee. A monochromatic X-ray beam with energy of 60 keV was used to image the breast cancer sample. Result of reconstruction is shown in figure 10(top). On this image, radiologists could easily identify the skin, fat and glandular tissue.

The training set is obtained from another breast sample imaged by the same technique

but with an high quality reconstruction. We consider a slice image, we apply a Sobel filter to extract the two derivative components and we run the KSVD algorithm.

Figure 9 shows the patches basis functions that we use to fit both components at the same time. The patches size is 7×7 pixels and each basis function is displayed as a 14×7 rectangle whose upper 7×7 part is the X component and the lower one the Y component.

Figure 10 is the reconstruction for a 765×765 pixels slice, using only 200 projections over the 1000 available. The upper left square is a zoom in the region marked in sub figure 10. The right column is the reconstruction with our method for X and Y components, while the left column is reconstructed with standard filtered back-projection using all 1000 available projections. Using our method we can still generate a high quality image with only one fifth of the projections which would be otherwise necessary to generate a high quality reconstruction with the standard FBP method. Visually the difference between the FBP results obtained with full data set and our method with a five-fold reduction of data is barely noticeable. The different borders of structures like skin layers, fatty tissues, and collagen strands are easily identified. The obtained result are very promising and a systematic evaluation for clinical application is under-way. The radiation dose absorbed by the sample during 200 projections is comparable to that of a standard clinical dual view (2D) mammography (3.5mGy).

For an eventual future clinical application of the PCI method it is important to investigate which is the acceptable compromise in terms of low dose and sufficient level of image quality. We need therefore to better explore how the quality of the reconstruction is degraded when we reduce the dose (i.e. number of projections and the acquisition time) further below the standard values. To this end, we performed a reconstruction with only 125 projections and results are shown in the figure III-D for one gradient differential image.

For an eventual future clinical application of the PCI method it is important to investigate which is the acceptable compromise in terms of low dose and sufficient level of image quality. We need therefore to better explore how the quality of the reconstruction is degraded when we reduce the dose (i.e. number of projections and the acquisition time) further below the standard values. To this end, we performed a reconstruction with only 125 projections and results are shown in the figure11. The first column present the result using our method, the second column is the result of reconstruction using FBP algorithm.

If a slightly higher noise level is tolerable, the method may be used with very few pro-

jections and thus applied to the screening and diagnosis of human breast cancers with an even lower radiation dose than conventional dual mammography. The results of our reconstruction show an image quality and a capability of discriminating fine structures that are still clinically acceptable. On the contrary, images produced with the standard FBP reconstruction method are very noisy and not diagnostically satisfactory.

IV. CONCLUSION

We have presented a new convex functional which implements in a mathematically pure form the concept of overlapping-patches-averaging, which was used so-far with a non-convex formalism. The resulting algorithm is robust, efficient, and well adapted to strongly reduce the noise in a natural image. The method was applied also to a medical diagnostic case by considering phase contrast tomographic data of whole cancer bearing human breasts acquired with phase contrast imaging. We demonstrated that it is possible to reduce the deposited dose in breast CT by a factor 5 compared to the standard algorithm while keeping the same image quality. Although we used this specific example as proof of principle in this study, the method we developed and described can be easily applied to other medical tomography fields. The numerical results have been generated with PyHST [11, 12], the ESRF tomography reconstruction code which uses the GPU implementation of the presented methods.

Acknowledgment

We thank Emmanuelle Gouillart and Gael Varoquaux for the interesting discussions and for having pointed us the possibilities of dictionary learning techniques and the examples contained in the scikit-learn package[13]. We thank the ESRF for providing the experimental facilities, the ESRF ID17 team for assistance in operating the facilities. We thank Alberto Bravin, Sergei Gasilov, Alberto Mittone for their help during experiments. This work was partially supported by the DFG-Cluster of Excellence Munich-Centre for Advanced Photon-

-
- [1] Bo Zhao, Huanjun Ding, Yang Lu, Ge Wang, Jun Zhao, and Sabee Molloy. Dual-dictionary learning-based iterative image reconstruction for spectral computed tomography application. *Physics in Medicine and Biology*, 57(24):8217, 2012.
 - [2] Patrick L. Combettes and Jean-Christophe Pesquet. Proximal Splitting Methods in Signal Processing. In H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, and H. Luke, D.R. adn Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, 2011.
 - [3] Francis Bach, Rodolphe Jenatton, Julien Mairal, and Guillaume Obozinski. *Optimization for Machine Learning*, chapter Convex Optimization with Sparsity-Inducing Norms. MIT Press, 2011.
 - [4] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, 2(1):183–202, March 2009.
 - [5] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, December 2006.
 - [6] Raazia Mazhar and Paul D. Gader. Ek-svd: Optimized dictionary design for sparse representations. In *ICPR*, pages 1–4. IEEE, 2008.
 - [7] Yu Hen Hu. Public domain images for homeworks and projects of the digital image processing course. <http://homepages.cae.wisc.edu/~ece533/>, 2012.
 - [8] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, April 2004.
 - [9] Alberto Bravin, Paola Coan, and Pekka Suortti. X-ray phase-contrast imaging: from pre-clinical applications towards clinics. *Physics in Medicine and Biology*, 58(1):R1, 2013.
 - [10] Yunzhe Zhao, Emmanuel Brun, Paola Coan, Zhifeng Huang, Aniko Sztrókay, Paul Claude Diemoz, Susanne Liebhardt, Alberto Mittone, Sergei Gasilov, Jianwei Miao, and Alberto Bravin. High-resolution, low-dose phase contrast X-ray tomography for 3D diagnosis of human breast cancers. *Proceedings of the National Academy of Sciences of the United States of*

- America*, 109(45):18290–18294, October 2012.
- [11] Alessandro Mirone, Emmanuelle Gouillart, Paul Tafforeau, Emmanuel Brun, and Jerome Kieffer. Pyhst2: an hybrid distributed code for high speed tomographic reconstruction with iterative reconstruction and a priori knowledge capabilities. *in preparation*, 2013.
 - [12] Alessandro Mirone. Gpl release of pyhst2 in preparation. <http://forge.epn-campus.eu/projects/pyhst2>, 2013.
 - [13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

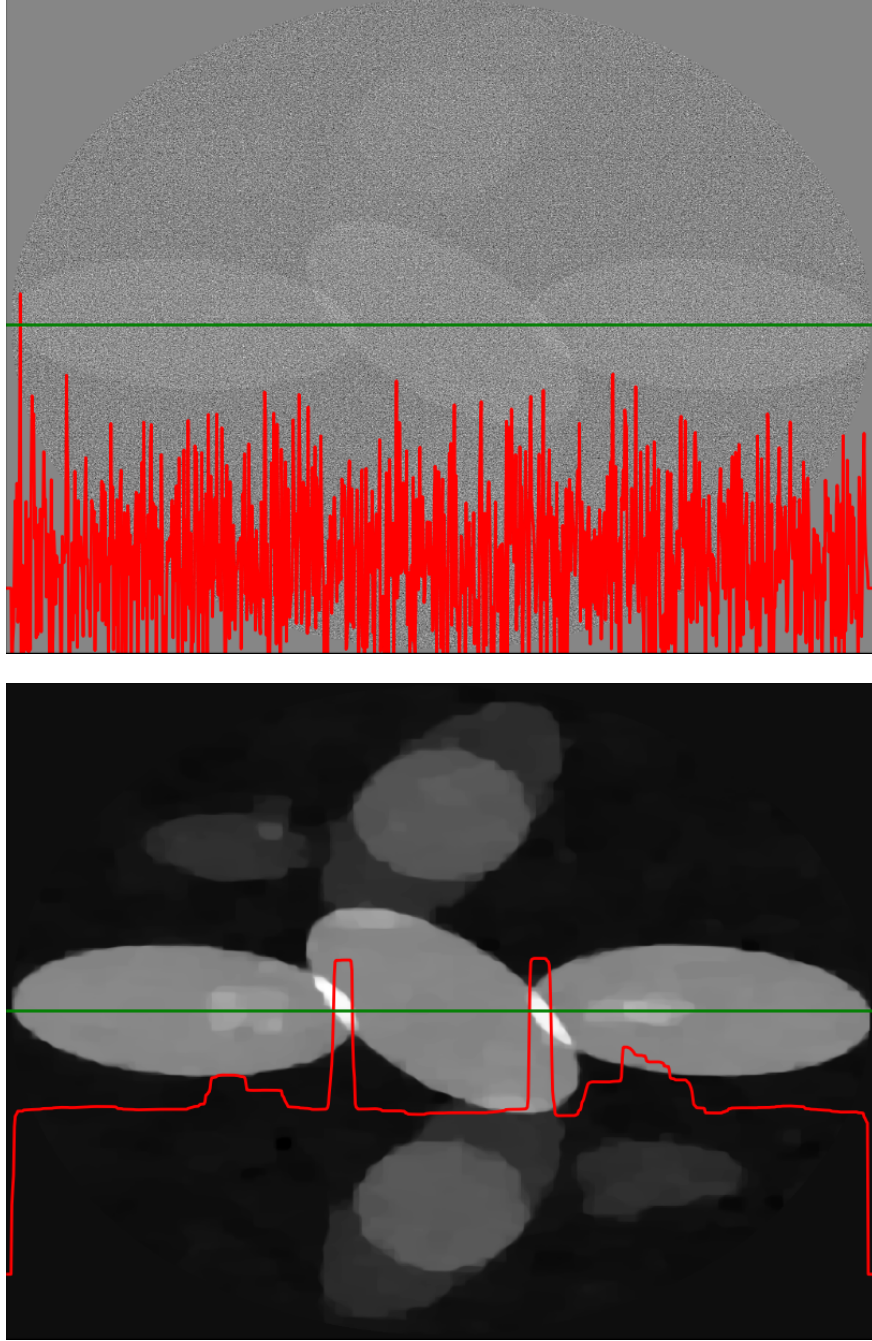


FIG. 8: Reconstruction from 150 angles of the phantom sinogram after adding a gaussian white noise with a σ equal to 5% of the maximum sinogram value. Top : FBP. Bottom: our method with $\beta = 1.1$, $\rho = 2.5 * 10^5$.

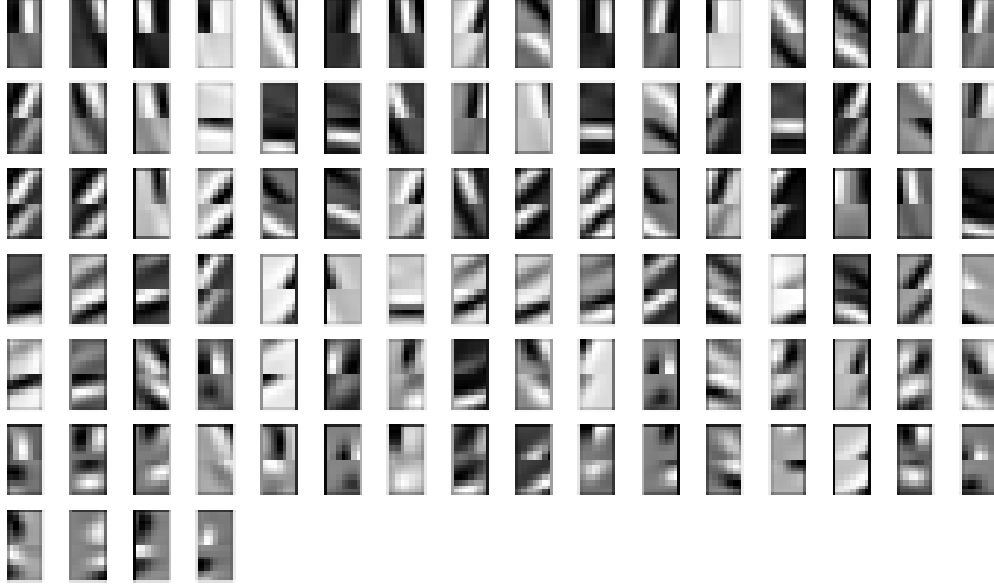


FIG. 9: The vectorial basis of patches learned from a high quality reconstruction of the phase gradient of a human breast. In each patch the upper 7×7 part is the X component while the lower part is the Y component.

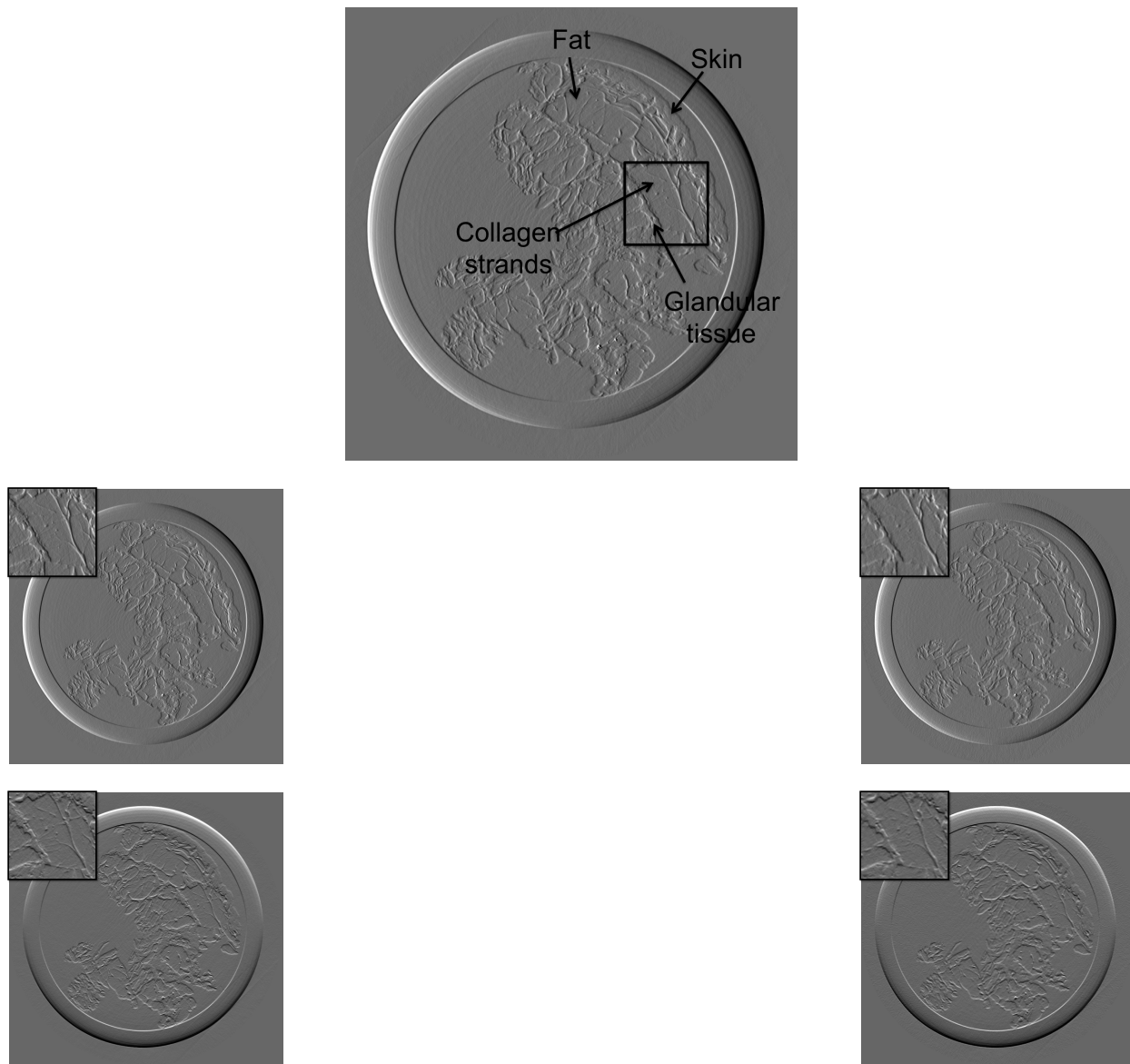


FIG. 10: Phase gradient X(first row) and Y(second row) components reconstruction. Left : FBP with the full set of data. Right: our method with one projection over five, using $\beta = 3 * 10^{-6}$ and $\rho = 10$.

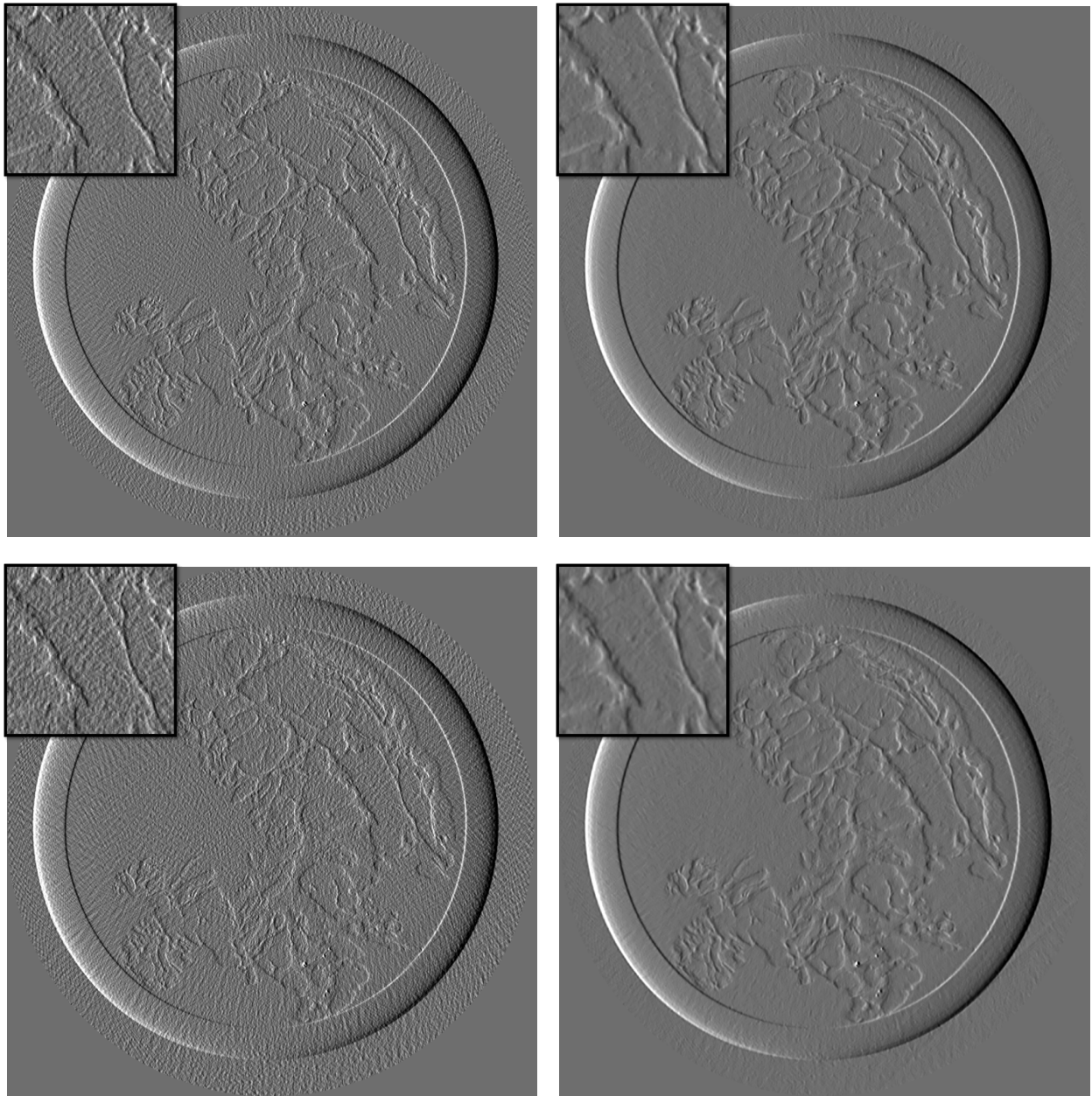


FIG. 11: Phase gradient X reconstruction. Left : FBP with 125 (top) and 100 projections. Right: our method using $\beta = 3 * 10^{-6}$ and $\rho = 10$.